PAG XXVIII
San Diego, CA - January 2020

Genomic predictions improve the performance of clonal cultivars in oil palm

Achille NYOUMA[1,2], Florence JACOB[3], Joseph BELL[1], Indra SYAHPUTRA[4], Dadang AFFANDI[4], Benoit COCHARD[3], Tristan DURAND-GASSELIN[3], David CROS[2,5,6]

[1]Faculty of Science, University of Yaounde I, Cameroon; [2]CETIC (African Center of Excellence in Information and Communication Technologies), University of Yaoundé 1, Cameroon; [3] PalmElit SAS, 34980 Montferrier sur Lez, France; [4]P.T. SOCFINDO Medan, Medan, Indonesia, [5]CIRAD (Centre de coopération Internationale en Recherche Agronomique pour le Développement), UMR AGAP, F-34398 Montpellier, France; [6]University of Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France

Abstract

Prediction of clonal genetic value is among the difficulties of the genetic improvement of oil palm (*Elaeis guineensis* Jacq.) yield. Presently, clonal selection requires two stages of phenotypic selection (PS): preselection on the phenotypic values of one or two yield components having high heritability, and final selection on performances in clonal trials. The current study evaluated the efficiency of genomic selection (GS) for clonal selection on eight traits. The GS models were trained on 295 and 279 Deli × La Mé crosses for bunch production and quality components, respectively, and were validated on 42 Deli × La Mé ortets of known clonal value. Genotyping by sequencing led to a dense genome coverage with 15,054 single nucleotide polymorphisms (SNP). We assessed the effects of SNP dataset (SNP density and quality) and of two GS modelling approaches on prediction accuracy. The results showed prediction accuracies that ranged between 0.70 and -0.03 according to trait, SNP dataset and model. Modeling disregarding the parental origin of alleles was preferable given the simplicity of implementation and the robustness over traits and SNP datasets, although including parental origin effects could slightly increase prediction accuracies for the traits used to define the two oil palm heterotic groups (bunch number and average bunch weight). The greatest GS prediction accuracies were beyond those of PS for most of the traits. Prediction accuracies from 0.70 to 0.45 for all traits can be achieved combining GS and PS. The best GS prediction accuracies are achieved with at least 7,000 SNPs. This will enable preselecting ortet candidates on all traits before clonal trials, thus increasing the selection intensity and the genetic progress.

Introduction

- Oil palm is the major oil crop worldwide with a production of **+65 Mt** (2018). The expected demand of palm oil in 2050 is **120-156 Mt**
- Clones (ramets) of top-ranking commercial hybrid individuals (ortets) can considerably increase the yield, hence the need of an efficient method to measure their total genetic value.
- Most yield components have a low heritability, therefore only one or two traits are considered for phenotypic preselection of clone candidates. Consequently, the estimation of the yield genetic values is of low reliability. For that reason, breeders set clonal trials. This gives an accurate genetic value estimation of yield but increases the selection process of 10 years and reduces the selection intensity.

Fig. 1. Oil palm

- **Objectives:** **evaluation of the efficiency of GS for clonal selection, using ortets of known clonal value to validate genomic predictions**

Materials and methods

Training set: 16 005 individuals

Validation set: 42 ortets and 3 558 ramets

Genotyping: **genotyping by sequencing (GBS) on the parents of the training crosses, 15,054 SNPs**

Dataset preparation: imputation and phasing

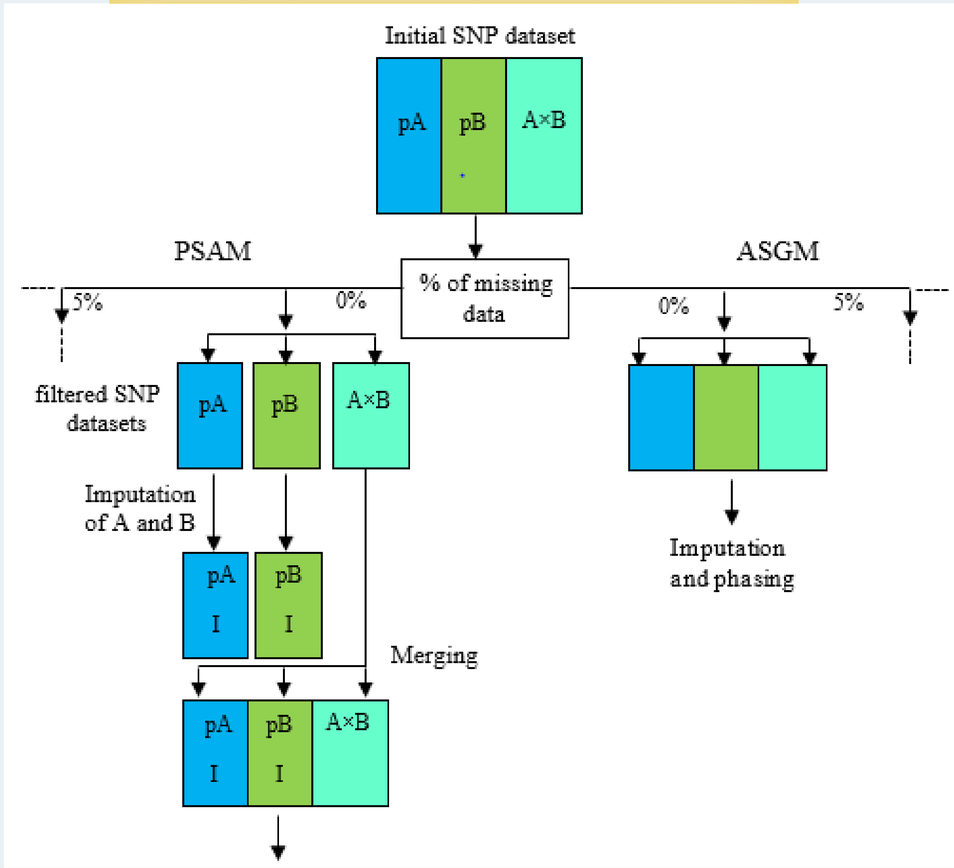

Fig. 2. Imputation and phasing scheme for the production of the SNP datasets used for genomic predictions with the two models PSAM (population-specific effects of SNP alleles model) and ASGM (across-population SNP genotype model)

Prediction models

❑ **Across-population SNP genotype models (ASGM)**

$$y = X\beta + Z_1 g_i + Z_2 g_{Deli \times LM} + Z_3 b + Z_4 p + \varepsilon$$

❑ **Population-specific effects SNP alleles models (PSAM)**

$$y = X\beta + Z_1 g_{Deli} + Z_2 g_{LM} + Z_3 g_{Deli \times LM} + Z_4 b + Z_5 p + \varepsilon$$

y: observed phenotypes of the training hybrid individuals, β: vector of fixed effects, g_i: individual additive genetic effects, $g_{Deli \times LM}$: dominance genetic effects, b: incomplete block effect and p: elementary plot effects. X, Z_1, Z_2, Z_3 and Z_4: incidence matrices and ε: vector of residual effects.

Results and Discussion

Results are presented for 3 traits /8

Effect of GS prediction model and SNP dataset on prediction accuracy

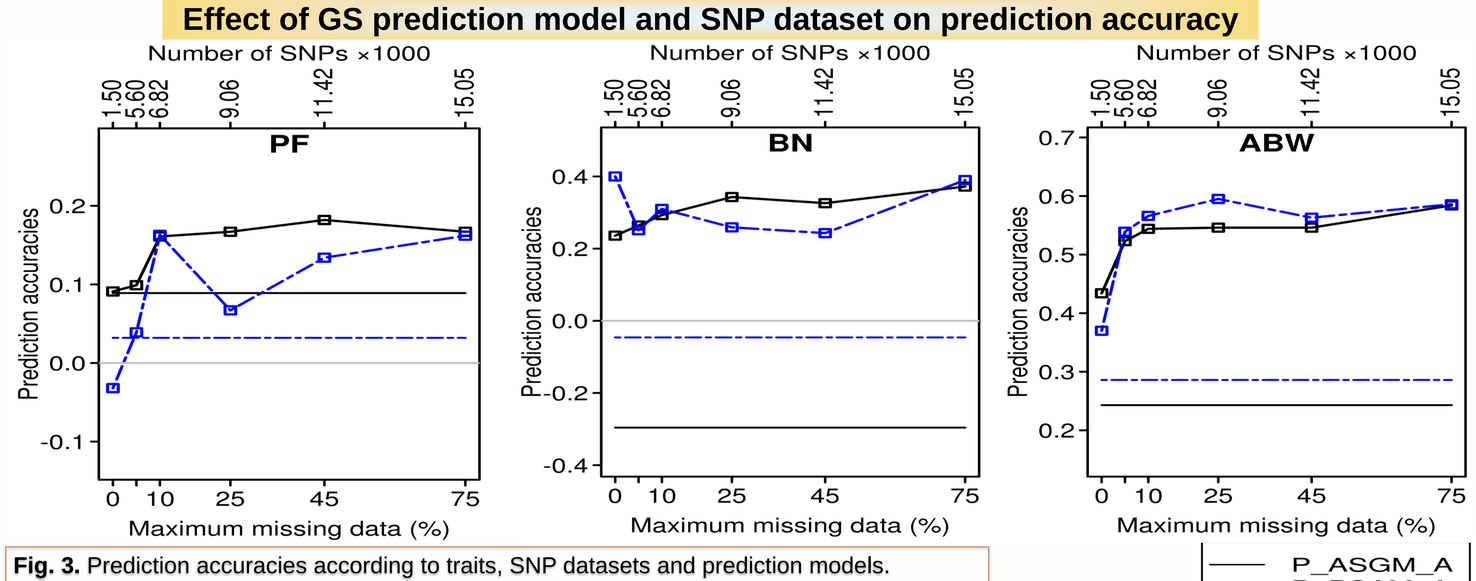

Fig. 3. Prediction accuracies according to traits, SNP datasets and prediction models.

G_ASGM_A: additive genomic across-population SNP genotype model, G_PSAM_A: additive genomic population-specific effects of SNP alleles model. **P_ASGM_A: additive pedigree-based across-population SNP genotype model, P_PSAM_A**: additive pedigree-based population-specific effects of SNP alleles model

Prediction accuracies of PS vs GS

GS > PS for 7 traits /8

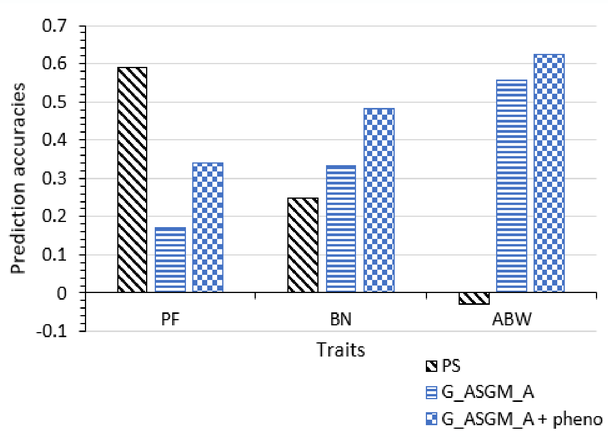

Fig. 4. Prediction accuracies of phenotypic selection and of the G_ASGM_A model without phenotypic data (G_ASGM_A) and with phenotypic data (G_ASGM_A+pheno) of ortets, on average over the best SNP datasets, and according to trait.

On average over all traits, the prediction accuracies of G_ASGM_A and G_ASGM_A+pheno were, respectively, 64.3% and 89.3% greater than PS.

- Annual genetic progress can be improved by selecting ortets (1) among a large population of the best possible crosses at the juvenile stage with GS models on most of the yield components or, (2) at the mature stage on all the yield components, using jointly the genomic and phenotypic data of the ortet selection candidates.
- Fst between Deli and La Mé around 0.55 (Cros et al., 2018), so G_PSAM_A was expected to perform better than G_ASGM_A on more traits. However, Zeng et al. (2013) attributed low accuracies in BSAM approach in animal studies to the complexity of the model caused by the segregation of SNP in both parental breeds. Also, G_PSAM_A is advantageous over G_ASGM_A on accuracy for a low marker density, a large number of records of the training set and a tightly relationship or absence of relationship between breeds (Ibánez Escriche et *al*., 2009, Technow et *al*., 2012, Lopes et *al*., 2017).
- Cros et al. (2017) obtained on hybrid crosses significantly higher prediction accuracies for OP, ABW, BN and FFB with GS than with the pedigree-based model.
- The genotypes of the phenotyped hybrid individuals constituting the training set were reconstructed using the molecular data of their parents, to make GS predictions cost effective. However, in the case of allele segregation distortion at a locus, the mean genotype in a hybrid family would significantly deviate from the mean genotype expected from the parental genotypes (Zuo et al., 2019; Li et al., 2015). Further studies should consider genotyping the training hybrid individuals.

Conclusion and Perspectives

- G_ASGM_A model provided suitable results when predicting clonal values for oil palm yield components (sufficient accuracy, robustness over traits and SNP datasets and easy implementation).
- Combining GS and PS enables efficient individual preselection before clonal trials over all the traits (with accuracies ranging from 0.45 to 0.70) instead of preselection on only one or two traits for the current PS.
- GS will improve the genetic gain in clonal cultivars.
- **Perspective**: Prediction of the genetic value of clones using genomic data of hybrids for calibration, with different imputation and phasing methods, and using a larger validation dataset.

Acknowledgments

The authors acknowledge SOCFINDO (Indonesia), CRAPP (Benin) and PalmElit for planning and carrying out the field trials with CIRAD (France) and authorizing use of the phenotypic data for this study. We thank Bertrand Pitollat (CIRAD) for help in cluster management and Nicolas Turnbull (PalmElit) for leaf sample collection in clonal trials. We acknowledge the CETIC for support, and we thank the UMR AGAP genotyping technology platform (CIRAD, Montpellier), the DArT company (www.diversityarrays.com) and the CIRAD-UMR AGAP HPC data center of the South Green bioinformatics platform (http://www.southgreen.fr/) for their help. This research was partly funded by a grant from PalmElit SAS.

References

- Cros, D., Bocs, S., Riou, V., Ortega-Abboud, E., Tisné, S., Argout, X., Pomiès, V., Nodichao, L., Lubis, Z., Cochard, B., 2017. Genomic preselection with genotyping-by-sequencing increases performance of commercial oil palm hybrid crosses. BMC genomics 18, 839. https://doi.org/10.1186/s12864-017-4179-3.
- Cros, D., Tchounke, B., Nkague-Nkamba, L., 2018. Training genomic selection models across several breeding cycles increases genetic gain in oil palm in silico study. Molecular Breeding 38, 89. https://doi.org/10.1007/s11032-018-0850-x.
- Ibánez-Escriche, N., Fernando, R., Toosi, A., Dekkers, J., 2009. Genomic selection of purebreds for crossbred performance. Genetics Selection Evolution 41, 12.
- Li, C., Bai, G., Chao, S., Wang, Z., 2015. A high-density SNP and SSR consensus map reveals segregation distortion regions in wheat. BioMed research international 2015.
- Lopes, M.S., Bovenhuis, H., Hidalgo, A.M., Van Arendonk, J.A., Knol, E.F., Bastiaansen, J.W., 2017. Genomic selection for crossbred performance accounting for breed-specific effects. Genetics Selection Evolution 49, 51.
- Technow, F., Riedelsheimer, C., Schrag, TobiasA., Melchinger, AlbrechtE., 2012. Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theor Appl Genet 125, 1181–1194. https://doi.org/10.1007/s00122-012-1905-8.
- Zuo, J.-F., Niu, Y., Cheng, P., Feng, J.-Y., Han, S.-F., Zhang, Y.-H., Shu, G., Wang, Y., Zhang, Y.-M., 2019. Effect of marker segregation distortion on high density linkage map construction and QTL mapping in Soybean (Glycine max L.). Heredity. https://doi.org/10.1038/s41437-019-0238-7.